

Claims

1. A document digitizing method of digitizing a document in printed form, comprising:

optically scanning the document;

forming and storing a digitized image file from the optically scanned document;

optically recognizing under computer control characters in the optically scanned document; and

forming and storing a text file of the optically recognized characters in document.

2. The method of claim 1 in which the document includes plural pages and a separate digitized image file is formed for each page of the document.

3. The method of claim 2 in which a separate text file is formed for each page of the document.

4. The method of claim 1 in which the document includes plural pages and a separate text file is formed for each page of the document.

5. The method of claim 1 in which each digitized image file is correlated with a corresponding text file.

6. The method of claim 5 in which corresponding digitized image files and text files are correlated by being assigned common names and are distinguished by appropriate file extensions.

7. The method of claim 5 in which corresponding digitized image files and text files are correlated by a mapping table or algorithm.

8. The method of claim 1 further comprising retrieving a digitized image

file for a document based upon a text string in the text file corresponding to the digitized image file.

9. The method of claim 1 in which the digitized image file is compressed and of a lossless image file format.

10. The method of claim 1 in which the text file is of a simplified file format based upon ASCII characters.

11. The method of claim 1 in which optical character recognition is applied to all text characters in the optically scanned document.

12. In a document digitizing system for optically scanning and digitizing a document in printed form and having a computer readable medium, a digitized document data structure for digitized documents stored in the computer readable, comprising:

a digitized image file representing a digitized image from the optically scanned document; and

a text file of text characters optically recognized from the optically scanned document.

13. The data structure of claim 12 further comprising correlated file indicators for the digitized image file and the text file representing the optically scanned document.

14. The data structure of claim 13 in which the correlated file indicators include common file names and distinct file extensions for the digitized image file and the text file representing the optically scanned document.

15. The data structure of claim 12 in which the document includes plural

pages, the data structure comprising a separate digitized image file for each page of the document and a corresponding separate text file for each page of the document.

16. The data structure of claim 12 in which the digitized image file is compressed and of a lossless image file format.

17. The data structure of claim 12 in which the text file is of a simplified file format based upon ASCII characters.

18. In a document digitizing system for optically scanning and forming a digitized image file of a document having text characters in printed form, a method of retrieving the digitized image file for a document, comprising:

storing digitized image files for plural printed documents in association with text files of the text characters in each document, the text files being generated by computer optical character recognition of the digitized image files or related image files;

searching the text files to identify any having a selected text string; and
providing access to the digitized image files corresponding to the text files identified as having the selected text string.

19. The method of claim 18 in which providing access to the digitized image files includes allowing a user to selectively display any of the digitized image files corresponding to the text files identified as having the selected text string.

20. The method of claim 18 in which digitized image files are associated with text files by having common file names and are distinguished by appropriate

file extensions.

21. The method of claim 18 in which searching the text files to identify any having a selected text string includes specifying multiple separate text strings and searching the text files in a batch to identify any text files having any of the separate text strings.

22. The method of claim 18 in which the text files have file names, the method further comprising storing the file names of the text files identified as having the selected text string.

23. In a computer-readable medium, document digitizing software for digitizing a document in printed form, comprising:

software for optically scanning the document;

software for forming and storing a digitized image file from the optically scanned document;

software for optically recognizing under computer control characters in the optically scanned document; and

software for forming and storing a text file of the optically recognized characters in document.

24. The medium of claim 23 further including software for correlating each digitized image file with a corresponding text file.

25. The medium of claim 24 in which corresponding digitized image files and text files are correlated by being assigned common names and are distinguished by appropriate file extensions.

26. The medium of claim 23 further comprising software for retrieving a

digitized image file for a document based upon a text string in the text file
corresponding to the digitized image file.